

## PATENT ABSTRACTS OF JAPAN

(11)Publication number : 09-186706

(43)Date of publication of application : 15.07.1997

(51)Int.Cl.

H04L 12/28

H04Q 3/00

(21)Application number : 08-331320

(71)Applicant : XEROX CORP

(22)Date of filing : 11.12.1996

(72)Inventor : LYLES JOSEPH B

(30)Priority

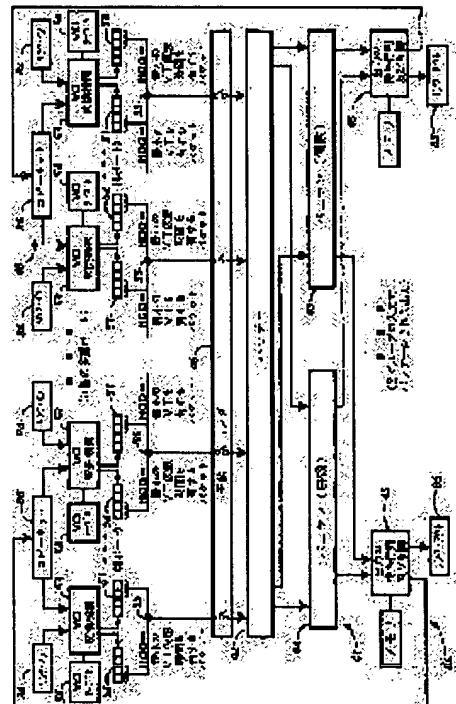
Priority number : 95 576099 Priority date : 21.12.1995 Priority country : US

**(54) METHOD FOR SOLVING COLLISION OF SCHEDULING BETWEEN PACKETS**

(57)Abstract:

**PROBLEM TO BE SOLVED:** To solve collision between inputs in contention to obtain an access with respect to output designation in a non-blocking exchange network routing a packet from plural inputs to plural outputs such as a high speed broad band communication network.

**SOLUTION:** An exchange is composed of a Batcher/Banyan network and includes a Batcher sorting network 42 and Banyan routing exchanges 43, 44. They are used to reduce output contention. In order to adjust the parallel processing, a buffer 45 is in existence to each output port to receive up to k-sets cells from the Banyan routing exchanges 43, 44. In order to judge impartially the contention among cells, a reservation link 46 is prepared, which solves the contention in existence among arbitration cycles and provides an access to each output in each arbitration cycle of contention parties.



## LEGAL STATUS

[Date of request for examination]

11.12.2003

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

Copyright (C); 1998,2003 Japan Patent Office

**THIS PAGE BLANK (USPTO)**

(19)日本国特許庁(JP)

(12) 公開特許公報(A)

(11)特許出願公開番号

特開平9-186706

(43)公開日 平成9年(1997)7月15日

(51)Int.Cl. <sup>6</sup>	識別記号	庁内整理番号	F I	技術表示箇所
H 0 4 L 12/28		9466-5K	H 0 4 L 11/20	G
H 0 4 Q 3/00			H 0 4 Q 3/00	

審査請求 未請求 請求項の数1 OL (全 20 頁)

(21)出願番号 特願平8-331320

(22)出願日 平成8年(1996)12月11日

(31)優先権主張番号 5 7 6 0 9 9

(32)優先日 1995年12月21日

(33)優先権主張国 米国(US)

(71)出願人 590000798

ゼロックス コーポレーション

XEROX CORPORATION

アメリカ合衆国 ニューヨーク州 14644

ロチェスター ゼロックス スクエア

(番地なし)

(72)発明者 ジョセフ・ビー・ライルズ

アメリカ合衆国 カリフォルニア州

94043 マウンテンビュー ローラレーン

2403

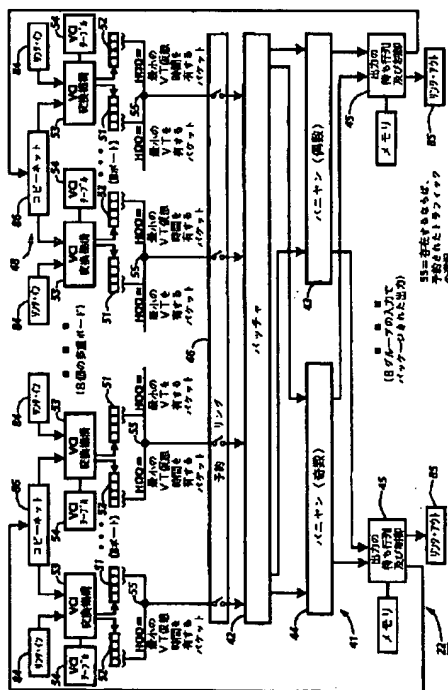
(74)代理人 弁理士 小堀 益 (外1名)

(54)【発明の名称】 パケットの間におけるスケジューリングの衝突を解決するための方法

(57)【要約】

【課題】 複数の入力から複数の出力へパケットをルーティングする非ブロッキング式交換網において、前記出力の指定のものに対するアクセスを求めて競合する入力の間における衝突を解決する予約リング機構を提供する。

【解決手段】 この予約リング機構は、少なくとも1つのアービトレーションサイクルにおいて頂部から底部へのリング状順序でステップのシーケンス及び比較オペレーションを実行する手段を含んでおり、自己刻時加重式の公正な待ち行列によって又は二者択一的に仮想時計によって要求される順序とこれもまた首尾一貫する頂部から底部への順序で前記指定の出力にアクセスする競合的な入力を許諾する。予め定められて許容された最大多数の競合者が、各々のアービトレーション・サイクルにおいて各々の出力に対するアクセスを付与される。



## 【特許請求の範囲】

【請求項1】  $k \geq 1$  である  $k$  個にまで達する異なった入力への所定等級のサービスに関するパケットを、前記入力におけるパケットのために潜在的に多数である入力によって作られたサービス要求に答えて、所定の出力へ同時にルーティングするように構成された有限帯域幅分散型多重化システムであって；各々の前記入力は、前記出力へパケットを供給するために前記帯域幅の予め定められた持ち分を割り当てられ、かつサービス許諾を留保されている前記パケットに関して少なくとも1つの記憶待ち行列を有するようにした前記多重化システムにおいて；前記パケットの間におけるスケジューリングの衝突を解決するための、次のステップを含む方法：前記入力によってそれらの受領の際に作られた各々のサービス要求に、(i) 要求を作る入力を一意に識別する識別子によって、更に(ii) 仮想終了時間によってラベル付けし、前記仮想終了時間は、要求を作る入力に割り当てられた帯域幅の持ち分に従って加重されるオフセットを加算することによって現行の仮想時間から計算されるようにし；ラベル付けされたサービス要求を分散型待ち行列に入力し；新しい要求が前記待ち行列へ入力される度に、それらのそれぞれの仮想終了時間に従って、ラベル付けされた要求を前記待ち行列内においてソーティングし、ラベル付けされた要求を仮想終了時間の昇順で組織化するようにし；前記多重化システムの各々のサイクルの間に、サービスに関する前記待ち行列から初めの  $k$  個までのラベル付けされた要求を取り出すようにする。

## 【発明の詳細な説明】

## 【0001】

【発明の属する技術分野】 本発明は、高速の広帯域通信網に関するものである。

## 【0002】

【従来の技術】 研究プロジェクト及び現今のシステム製品は、既存のコンピュータネットワークの能力を超過する帯域幅、応答性、及び／又はリアルタイムに関する要求を有するものである。例えば、個人的な用紙から技術的なライブラリーにまで及ぶ文書に関する紙から電子フォーマットへという進行しつつある移行は、進歩的な電子的走査装置及び印刷装置の開発に至っており、これらの装置は、これらの装置の速度及び解像度が増大するにつれて、加えてこれらの装置のカラー及びグレースケール能力が拡大されるにつれて、増大し続けるであろう帯域幅の要求を有する。

【0003】 多くのアプリケーションは、それらにどのような帯域幅が与えられるのか、それらが獲得し得た帯域幅をどの程度有益に利用し得るのかに応じて存続が可能である。これらのアプリケーションは弾力的な帯域幅要求を有するものである。コンピュータネットワークにおいて現在運用されているアプリケーションの殆どは弾力的な帯域幅要求を有するのである。これらのアプリケ

ーションを維持するサービス等級(class)は、インターネット・コミュニティにおける「ベスト・エフォート(best effort)」及び広帯域ISDN/ATMコミュニティにおける利用可能なビット速度(ABR)として知られている。帯域幅がこのサービス等級のアプリケーションの間で「公正に(fairly)」分割されないと、様々の不都合な現象が生じる可能性があることが知られている。

【0004】 非同期伝達モード(ATM)交換は、現在、多数のサービス等級を維持するだけでなく前述のアプリケーションをも維持するに足る十分なピーク帯域幅及び統合帯域幅を提供するものとして受け入れられた機構になりつつある。多くのATM交換器は、現在、FIFO待ち行列によって実現されている。FIFO待ち行列は、ABRトラフィックのために使用されるとき異常な行動を示すものである(「パケット交換式ゲートウェイにおけるオン・トラフィック位相効果(On Traffic Phase Effects in Packet-Switched Gateways)」サリー・フロイド及びヴァン・ヤコブソン(Sally Floyd and Van Jacobson)著、インターネットワーキング：研究と体験(Internet working: Research and Experience)、第3巻115から156頁(1992年)、及び「輻輳制御アルゴリズムの動力学における観察：二方向トラフィックの影響(Observations on the Dynamics of a Congestion Control Algorithm: The effects of Two-Way Traffic)」リシア・チャン、スコット・シェンカー及びデビッド・クラーク(Lixia Zhang, Scott Shenker, and David Clark)著、ACMシグコム91協議会(ACM Sigcomm 91 Conference)、9月3-6、1991年、スイス、チューリッヒ(Zurich, Switzerland)、133から148頁を参照すること)。FIFOは、正しく行動するユーザーを異常に行動するユーザーから保護することも不可能である(それは隔離を提供しない)。これらの欠陥の結果として、加重される公正な待ち行列のような非FIFO待ち行列機構(例えば、A. デマーズ、S. ケシェイブ及びS. シェンカー(A. Demers, S. Keshave, and S. Shenker)著、「公正な待ち行列アルゴリズムの分析及びシミュレーション(Analysis and Simulation of a Fair Queueing Algorithm)」、ACMシグコムの会報における1から12頁、1989年9月、及びA. K. パレク(A. K. Parekh)著、「統合サービスネットワークにおけるフロー制御に対する一般

化プロセッサ・シェアリング・アプローチ (A Generalized Processor Sharing Approach to Flow Control in Integrated Service Networks)」博士論文、電子工学及びコンピュータ科学部門、MIT、1992年を参照すること)、或いはラウンドロビンのような公正な待ち行列への近似化 (エレン・L. ハーネ (Ellen L. Hahn) 著、「データネットワークにおける最大-最小の公正さに関するラウンドロビン・スケジューリング (Round-robin Scheduling for Max-Min Fairness in Data Networks)」、通信の選択エリアにおけるIEEEジャーナル (IEEE Journal on Selected Areas in Communications)、第9巻1024から1039頁、1991年9月) が、しばしば提案されている。

【0005】もう1つのサービス等級 (即ちサービス等級の集合) が存在する: 回路エミュレーション及びビデオのようなリアルタイムのアプリケーションを維持するものである。これらのアプリケーションは、データが限定的ジッタ (即ちパケット遅延変動) によってネットワークを通じて送信されることを要求するかもしれない。A. パレクによる先に引用した文書で示されていたように、加重される公正な待ち行列は、リアルタイムストリームに有界ジッタを提供するために使用されることが可能である。パレクの成果は、近頃になって (ポワン・ゴヤル、サイモン・S. ラム及びハリック・M. ヴァン (Pawan Goyal, Simon S. Lam and Harrick M. Vin) 著、「異質のネットワークにおける端末間遅延限界の決定 (Determining End-to-End Delay Bounds in Heterogeneous Networks)」、デジタルオーディオ及びビデオのためのネットワーク及びオペレーティングシステムサポート (NOSSDAV) に関する第5回国際ワークショップの会報において、ニューハンプシャー、ダーラム (Durham, N. H.)、4月18-22、1995年)、仮想時計 (リシヤ・チャン著、「仮想時計: パケット交換網のための新しいトラフィック制御アルゴリズム (Virtual Clock: A New Traffic Control Algorithm for Packet Switching Networks)」、ACMシグコンの会報における19から29頁、1990年8月) 及び自己刻時式の公正な待ち行列 (S. J. ゴールスタニ (S. J. Golestani) 著、「高速アプリケーションのための自己刻時式の公正な待ち行列機構 (A Self-Clocked Fair Queuing Scheme for High Speed Application

s)」、インフォコム (INFOCOM) の会報における636から646頁、1994年) に密接に関連する機構を使用するシステムのための遅延境界を証明するように拡張された。

#### 【0006】

【発明が解決しようとする課題】従って、弾力的な両者 (ベスト・エファット/ABR) 及び非弾力的な (即ちリアルタイム) サービスは、公正な待ち行列及び関連するアルゴリズムの使用から利益を得ることが可能である。しかし、そのような機構は、待ち行列が発生する各々の多重化ポイントにおいて実現されなければならない。

【0007】公正な待ち行列及び関連するアルゴリズムは、パケットのシーケンスにおいて機能する (ATMセルがこの議論のためのパケットである)。ATMの場合、これらのシーケンスは、VCI又はVPIのいずれかによって識別され、一方、インターネットのプロトコルの組では、識別は、<IPアドレス、プロトコル、ポート>3つの部分から成る (IPv4) 又はフロー識別子 (IPv6) を根拠としている。自己刻時加重式の公正な待ち行列及び仮想時計の両者において、パケットは、タイムスタンプによって順序付け (ソート) される (ラウンドロビンのような機構はタイムスタンプによるパケットの順序付けへの近似化を提供する)。これらのタイムスタンプは、パケットに関する仮想終了時間 (virtual finishing time) を表現するものであり、開始時間値を取って、固有のパケットシーケンスの帯域幅の持ち分を表現する加重をパケットの長さに乗算することによって得られるオフセットを加算することによって計算される。

【0008】仮想時計の場合、仮想終了時間は、次のように計算される:

$$VT(f, 0) = 0$$

$$VT(f, J+1) = \max \{ \text{到着}(f, J+1), VT(f, J) \} + \text{長さ}(f, J+1) / \text{速度}(f)$$

ここで、 $VT(f, J)$  はフロー (仮想回路)  $f$  のパケット  $J$  に関連する仮想終了時間、到着 ( $f, J$ ) はフロー  $f$  のパケット  $J$  の到着時間、長さ ( $f, J$ ) はフロー  $f$  のパケット  $J$  の長さである。自己刻時加重式の公正な待ち行列は、次の公式に従って仮想終了時間を割り当てる:

$$VT(f, 0) = 0$$

$$VT(f, J+1) = \max \{ \text{システム仮想時間}, VT(f, J) \} + \text{長さ}(f, J+1) * \text{加重}(f)$$

ここで、システム仮想時間は、パケット ( $f, J+1$ ) が到着する時点でサービス (出力) されるパケットに関連する仮想時間であると規定される。ATMの場合、パケット長さは、セルが一定寸法のものであり、従って両者の式における最も右側の項がフローについて一定であるので、不変である。仮想時計の場合、簡略化された式

は次の通りである：

$VT(f, J+1) = \max \{ \text{到着}(f, J+1), VT(f, J) \} + \text{定数}(f)$

そして、自己刻時加重式の公正な待ち行列も類似の簡略化を有する。結果として、仮想時計又は自己刻時加重式の公正な待ち行列のいずれかを実現するATM待ち行列ポイントは、以下のステップを実行しなければならない：

1) VCに関する現行の仮想時間の最大値と、a) セルの到着時間か又はb) システム仮想時間のいずれかを計算せよ。

2) ステップ1の結果に、帯域幅のVCの持ち分を表現するVCについての定数を加算せよ。

3) ステップ1及び2によって割り当てられた仮想のタイムスタンプの増大する値の順序でセルをサービス（それらを送信）せよ。

【0009】

【課題を解決するための手段】本発明によれば、複数の入力から複数の出力へパケットをルーティングする非ブロッキング式交換網は、前記出力の指定のものに対するアクセスを求めて競合する入力の間における衝突を解決する予約リング機構を包含する。この予約リング機構は、少なくとも1つのアービトレーションサイクルにおいて頂部から底部へのリング状順序でステップのシーケンス及び比較オペレーションを実行する手段を含んでおり、自己刻時加重式の公正な待ち行列によって又は二者択一的に仮想時計によって要求される順序とこれもまた首尾一貫する頂部から底部への順序で前記指定の出力にアクセスする競合的な入力を許諾する。予め定められて許容された最大多数の競合者が、各々のアービトレーション・サイクルにおいて各々の出力に対するアクセスを付与される。

【0010】

【発明の実施の形態】図1は、階層的なATMネットワークを概略的に示している。図2は、ATM交換器を概略的に示している。図3Aは、ATMセルのための通常のフォーマットを示している。図3Bは、ATMセルの典型的な交換器ヘッダを示している。図4は、バッチャ・ソーティング要素を概略的に示している。図5は、バッチャ・ソーティング要素のオペレーションを示す有限状態機械である。図6は、図4で示された形式のソーティング要素の再帰的な組合せから構成されるバッチャ・ソーティング・ネットワークを概略的に示している。図7は、バニヤン・ルーティング・ネットワークを概略的に示している。図8は、本発明を実行するものとして適切である予約リングの簡略化された概略図である。

【0011】I. 定義

本文において使用される以下の用語は、以下の意味を有する：「チャンネル速度(channel rate)」は、例えば、単一のテレビジョン送信、ファイル転送、

データベーストランザクションのような特定のストリーム、チャンネルなどのビット速度である。「リンク速度(link rate)」は、ネットワーク装置(ホスト、ルータ、交換器)が個々のリンク(ワイヤ対、同軸ケーブル、光ファイバー)に関して持続可能であるか又は持続しなければならないビット速度である。この速度は、チャンネル速度の上限である。それは、インターフェース・ハードウェアのコスト、及びネットワークプロトコルのハードウェア及びソフトウェアのコストに対して主要な影響を有するものでもある。「統合速度(aggregate rate)」は、同時に送信され得るリンクの最大数に関するリンク速度の和として表現される最大の総計ネットワーク能力である。バス又はリングとして実現されるか又は単一周波数の無線放送を使用するネットワークの場合、リンク速度は、統合速度と同一である。逆に、従来型の電話交換システムは、いかなるリンクの速度よりも遥かに高い統合速度を提供する。

【0012】II. 基本的アーキテクチャ

本発明のATM交換器は、様々なスイッチング構造、中でもVLSIベースの「バッチャ/バニヤン(Batcher/banyan)」スイッチング構造、自動ルーティング・クロスバー、或いは補足的なロジックユニットを備えた既製のクロスバーを使用して実現されることが可能である。後続の議論では、バッチャ/バニヤン構造が想定されることになる。

【0013】ATM交換のアピールの1つは、実質的な標準化が起こったことである。この形式のネットワークの基本的なコンセプトは、すべての通信が小型で一定寸法の「セル」と呼ばれるデータパケットの送信及び交換によって成し遂げられることである。各々のパケットのオーバーヘッドは、仮想回路技術を使用することによって低減され、それにより、各々のセルのヘッダは、前もって供給元と宛先の間のすべての交換器を介して設定された通路を識別し、セルとその所望のルーティングを識別するために要求される情報が、従来型のXNS又はTCP/IPデータグラムにおけるヘッダ情報よりも遥かに小さくなる。

【0014】開発されたATM標準に従えば、セルは、そのヘッダが、宛先仮想回路インデックス、VCIとして知られる宛先アドレス情報を包含する小型で一定寸法(53バイト)のパケットである。仮想通路インデックスも存在するが、仮想通路における交換は、仮想回路における交換と本質的に同一であるので、その主題が本文において更に議論される必要はない。セルがスイッチング構造に入る前に、供給元ホストは、「仮想回路」として知られた固定ルートがセルをその供給元からその宛先へ送信するものとして決定されるようにする。この仮想回路内における各々の交換リンク(交換器から交換器への各々のリンク及び宛先リンクへの各々の交換器)は、交換器の固有仮想回路識別子(VCI')によって一意

に識別される。従って、1つの仮想回路は、多数の名称を有する。セルは、仮想回路に沿ったいかなるポイントにおいても再度順序付けされないかもしれないので、セルは送られた順序で到着しなければならない。

【0015】前述のことを心に留めて図1を参照すると、ATMセル27のようなセルを1つ又はそれ以上の供給元31及び32から1つ又はそれ以上の宛先33へ送信する多数の自動ルーティング、非ブロッキング式交換器22-25から構成される階層的なネットワーク21が存在することが理解されるであろう。例えば、LAN（ローカル・エリア・ネットワーク）の場合、供給元31及び32はワークステーションであるかもし、一方、宛先33はファイルサーバであるかもしれない。図示されたように、ネットワーク21は2レベルの階層を有するが、その階層が、メトロポリタン・エリア・ネットワーク（MAN）及びワイド・エリア・ネットワーク（WAN）さえも包含する遥かに長い通路を介するパケット搬送及び交換を提供するように拡張され得ることは明白であろう。

【0016】図2を参照すると、交換器22-25は、適切には、それぞれバッチャ／パニヤン・ネットワーク41から構成される。交換器22に関して図示されたように、単一nポートのバッチャ・ソーティング・ネットワーク42は、k個と同じほど多くのセルが同じセルの交換サイクルの間にそれらのセルの中のいずれのものも失うことなく同じ出力ポート番号にルーティングされ得るようにして、接続されるk個（典型的には2又は3）のパニヤン・ルーティング交換器43及び44を供給する。この並行処理を調整するため、各々の出力ポートには比較的大きな（数千個のセル）バッファ45が存在し、このバッファは、出力パニヤン43及び44からサイクル毎にk個まで達するセルを受け入れるように構成される。リアルタイム・トラフィック（確定的ビット速度、即ちDBR）及び「ベスト・エファット」（利用可能なビット速度、即ちABR）トラフィックの両者を処理するため、バッファ45は、異なる寸法の2つ又はそれ以上のサブ・バッファ（図示略）から有益に構成され：速度が限定されるか又は予約されたトラフィックのための数百セルの小さなバッファと、予約されないトラフィックのための遥かに大きなバッファとを包含する。数学的な成果は、加重される公正な待ち行列交換器の場合には、VCによって経験される最大遅延が、従って、そのVCのトラフィックのために要求され、かつ深さb及び速度rのトークン・パケットに適合するように形成される最大バッファ割当てもまた、bとサービス速度の関数であることを示している。交換ノードにおけるバッファ割当ての総計は、そのとき、VCの各々のためのバッファ要求の和である。パケット寸法がセル遅延変動（CDV）を決定し、DBRサービス等級が小さなセル遅延変動を予測するので、そのとき（関連する速度のた

めに）これはDBRトラフィックのためのバッファが比較的小さいものでなければならないことを暗示する（例えば、各々の能動VCについて1つ又は2つのセルというオーダー）。厳密な速度限界で限定されない現行のデータネットワークで発生するような「爆発的なトラフィック」の場合、関連する仮想回路は、大きなバッファ（ベスト・エファット／ABRトラフィック）において待ち行列に入れられる。ABRサービスのための制御ループは、緩衝に値する少なくとも1回の往復を必要とする。単一の出力ポートの上を走るABRのVCに割り当てられた帯域幅の和が出力ポートのリンク速度より大きくなることは決してあり得ないので、要求される緩衝の総計は、リンク速度掛ける最大往復時間に比例する。北米大陸の合衆国内におけるサービスの場合、これは約30ミリ秒であり、従って150メガビット／秒の交換器は、出力線について凡そ1/2メガバイトの緩衝を必要とする。周知のように、これらの寸法のバッファは、SRAMから容易に構築される。より速い速度及びより長い往復時間は、同期DRAM（SDRAM）のような先進的なDRAM技術を使用することによって、或いは緩衝記憶装置へのデータ通路の並行処理を増大させることによって容易に調整される。

【0017】理解されるであろうが、k個以上のセルが単一の交換器サイクルの間に同じ出力ポート番号にルーティングされるとしたら、交換器は過剰なセルを単純に脱落させるのではなく、期待されたようには機能できないことになる。それ故、この状況は、回避されなければならない。そのため、受け入れられている慣例に従って、それらの中の最大でもk個のものがいずれのサイクルにおいても交換器に提供されることを許容しつつ、そのように衝突するセルの間の競合を公正に裁く予約リング46が準備される。

【0018】スイッチング構造41の入力側へ接続され、そこで予約リング46へ適切に接続されるのは、既製品又は注文品のいずれかのクロスバー交換器構成要素を使用して構成され得るn×nのコピーネットワーク48である。すべてのセルは、ポイント・トゥー・ポイント方式で、或いはそれらの交換器ヘッダにおけるマルチキャスト・ビットの状態によってマルチキャスト・セルとして詳細にマークされる。理解されるであろうが、交換器ヘッダは、ATMセルに対して前部に付録され、スイッチング構造41を介してそれらのセルをルーティングさせ、参照番号45の所でそれらを適切な出力待ち行列へ入れる。図示された実施例において、入ってくるマルチキャスト・セルは、ユニキャスト・セルと全く同じようにスイッチング構造41をルーティングされるが、その後、出線ルーティングされるのではなく、スイッチング構造の出力において中断され、コピーネットワーク48へ戻される。

【0019】更に、交換器41の入力ポートの各々には

記憶装置50も存在し、入ってくるセルとコピーネットワーク48によって供給されたセルのそれぞれを緩衝する。この記憶装置は、入力ポートに到着したセルを競合なしで交換器41を介してルーティングされ得るまで保持する2つのバッファ51及び52に分割される(所望されるならば、これらのバッファの1つ又は両方は、多数の優先権集合を支持するためにサブ・バッファに細分されても良い)。所望されるならば、緩衝記憶装置50は、よりコンパクトな実行を達成するために、VCI変換機構参照用テーブル54と組み合わせられることも可能である。

【0020】交換ネットワーク41を介するセルのルーティングは、交換器の入力ポートにおいてVCI変換機構53によって制御される。これらの変換機構53は、関連する入力ポートのための待ち行列の先頭においてセルのヘッダによって担持される宛先仮想回路インデックス(VCI)及び入力ポートアドレスの3値関数、即ち、[VCI<sub>i</sub>、出力ポート、マルチキャストフラグ]ルーティング関数[iVCI]を計算する、ホスト・プロセッサ(図示略)によって維持され更新されるVCIテーブル参照用記憶装置54を使用する。認識されるであろうが、従来型のマイクロプロセッサなど(図示略)へのインターフェースは、主として様々なルーティング及びマッピング・テーブルをそのVCI参照用テーブル54へロードすることによって、交換器22を制御するために要求される。

【0021】III. 交換器の構成要素に関する更に詳細な説明

#### A. 仮想回路変換機構

スイッチング構造41の入力ポートにおけるアドレス・マッピングを実行するために準備される変換機構53は、以下のものを包含する幾つかの機能を有している：

所望の宛先へのルートの途上にある構造41から出力ポートを(入力待ち行列の先頭においてATMセルのヘッダに包含される宛先VCI/VPIDアドレスに基づいて)選択すること；入ってくるVCI/VPIDアドレスを出リンクのための適切な値に書き直すこと(思い出されるであろうが、VCI/VPIDのものは端末間の重要性を有するものではない)；及び、以下でより詳細に説明されるように、マルチキャスト送信のために要求される特殊なマッピングを処理することである。これらの変換機構は、ビルド及びロード監視目的などのために仮想回路について交換されるATMセルの個数の計数を維持することも可能である。図3Aは典型的なATMセルのフォーマットを示し、図3Bはそのようなセルのための典型的な交換器ヘッダのフォーマットを示す。セルヘッダにおける大域的フロー制御(GFC)フィールドは、通常、空白の値を包含していることが留意されるべきである。

【0022】B. バッチャ/パニヤン・ネットワーク

スイッチング・ネットワークは、いかなる交換サイクルにおいても、その入力に関して任意順列又は部分的な順列を生成することが可能であるならば、「非ブロッキング」であると呼ばれる。言い換えれば、非ブロッキング式交換器は、単一の交換サイクルにおいてすべての可能な1対1の入力/出力マッピングを生成することが可能なのである。

【0023】スイッチング・ネットワークは、典型的には、単純な交換器又はステージの階段状ネットワークから構成される。そのようなネットワークを介するルートは、所定の入力から所定の出力への交換器ステージを横断する通路である。スイッチング・ネットワークは、ネットにおける各々のステージに関して、ルーティングが、交換器のステージへ進入するセルの中に包含された情報のみによって、即ち位置的に利用可能な情報によって決定され得るならば、「自動ルーティング」であると分類される。

【0024】バッチャ/パニヤン・ネットワークは、非ブロッキング、自動ルーティング式スイッチング構造の既知の具体例である。周知のように、バッチャ・ネットワークは、典型的には深さ $(\log_2 N)^2$ のネットワークによってNデータストリームを分類する並行ソーティング・ネットワークであり、このことは、 $N(\log_2 N)^2$ のソーティング要素が要求されることを意味する。図4で示されたように、各々のソーティング要素62は、2つの入力(「A」及び「B」)を受け入れ、それらの最小値と最大値の両方を計算し、それによってそれらの入力を分類する。従って、入力が最も重要なビットを初めに備えた2ビット連続データストリームの形態であるとき、ソーティング要素は、図5の参照番号64のように、非常に簡単な有限状態機械として機能する。これらのソーティング要素62は、従来的には、4個、8個、又はそれ以上の入力のためのソーターを構築するために、再帰的に組み合わせられている(図6を参照)。

【0025】図7で示されたように、標準的なパニヤン・ルーティング・ネットワーク65は、深さ $\log_2 N$ のマルチステージ・ネットワークであり、ここで、Nはネットワークに対する入力の個数である。バッチャ・ソーティング・ネットワーク65と同様に、パニヤン・ルーティング・ネットワーク71は、各々が2つの入力と2つの出力を有する基本的なスイッチング要素72から再帰的に構成される。従って、パニヤン・ネットワーク71には $O(N \log N)$ の基本的なスイッチング要素72が存在する。データシーケンスがパニヤン・ネットワーク71を介して流れるとき、ネットワーク71の各々のステージにおける基本的要素72は、出力アドレスの1つのビットを検査し、そのビット及びその特定のシーケンスにおけるすべての後続のビットを、そのビットの値に依存して、その出力の1つ又はその他にルーティングする。標準的な慣習に従って、アドレスビットは各



々のステージに関して変化し、それによって、第1のステージは最高順位のビットの値に基づき、第2のステージは2番目に高い順位のアドレスビットの値に基づき、N番目のステージは最下位の順位のビットの値に基づいてルーティングする。N個のアドレスビットが処理された後、シーケンスの残りのものは、交換器出力に対して確立された通路をたどることになる。これは、1つのシーケンスの終了と次のものの始まりが外部要因によって決定されなければならないことを意味する。従って、この事例では、各々のシーケンスは一定寸法のATMセル・プラスその一定寸法の交換器ヘッダであることが思い出されるべきである。

【0026】パニヤン・ネットワーク71のいかなるステージにおいても、両方の入力と同じ出力を選択することが可能である。それが生じると、パニヤン・ネットワークは「ブロック」するので、それは不確定な結果を産み出す。言い換えれば、パニヤン・ネットワーク71は、その入力の任意順列を計算することができないが、順列の集合が存在し、それ故に、それは「非ブロッキング」なのである。これらの集合の中の1つは、入力10 がそれらの出力によって順序付けされるものである。その理由のため、図2で示されたように、バッチャ・ソーティング・ネットワーク42はパニヤン・ルーティング・ネットワーク43及び44に先行し、それによって、非ブロッキング式かつ自動ルーティング式であるスイッチング構造41を提供するのである。

【0027】図2で示されたように、複数のパニヤン43及び44（典型的には2つ又は3つ）は、出力競合を低減させるために有益に使用される。例えば、図示の実施例では、バッチャ・ソーティング・ネットワーク42の後に続く2つのパニヤン・スイッチング・ネットワーク43及び44が存在し、 $k=2$ のスピードアップ要因を提供する。パニヤン43及び44の各々はその入力の $1/k$ のみを使用するので、その他の入力は途中で切り上げられる。バッチャ42の出力は、続いて、パニヤン43及び44の入力に接続されるので、バッチャ42のk番目毎の出力はパニヤン43又はパニヤン44のk番目の入力に接続される（示されたように、これは、パニヤン43及び44がバッチャ42の奇数及び偶数に番号付けされた出力をそれぞれにルーティングすることを意味する）。結果として、同じ出力ポートにアドレスされたk個までの明確なセルは、バッチャ42のk個の連続的な出力において出現することが可能であり、その後、k個の明確なパニヤン43及び44を介してスイッチング構造41の所望の出力ポートにルーティングされ得るのである。

#### 【0028】C. 予約リング

予約リング46（図2）は、単一の出力ポート宛のセルの個数が「k」を超過するとき、スイッチング構造41の出力ポートへの「公正な」アクセスを実現しつつ、出

力競合を解決するアービタである。「公正さ」には多くの異なった定義が存在するが、自己刻時加重式の公正な待ち行列によって指定された順序でスイッチング構造にセルを入れるアービタは、弾力的なサービスを支持するトラフィック集合のために好適であり、更に自己刻時加重式の公正な待ち行列又は仮想時計のいずれかが、リアルタイムのサービスを支持するトラフィック集合のために好適である。

【0029】より詳細には、予約リング46は、リング状の形状に互いに接続される有限状態機械76a-76nの線形収縮配列（linear systolic array）75（図8）として実現される。本文では以下において「評価機構」と呼ばれることもあるこれらの有限状態機械の各々は、スイッチング構造41の入力ポートのそれぞれの1つと連絡する。理解されるであろうが、この予約リングは、同じ明快な優先権を備え、所定のアービトレーション・セッションの間に同じ出力宛先に仕向けられていると判明した、すべてのセル即ち「パケット」に、その所定のセッションに参加しているセルのすべてがそれらの共通の出力宛先へルーティングされてしまうまで後続の到着分に対する閉集合を適切に形成させる。これは、各々の出力宛先に関する各々のアービトレーション・セッションが、その同じ宛先に関するもう1つのセッションが開始される前に、順序正しく決定されることを保証する。配列75のような線形収縮配列を使用することの利点の中でもとりわけ、すべての連絡が局所的であって、電気的な負荷が配列の寸法と共に大きくならないことは有益である。これは、システムの線形スケールリングを可能にする。

【0030】予約リングは、複数のルーティング・ネットワーク43及び44（即ち、スピードアップ要因、 $k \geq 2$ を有するネットワーク）を有する交換器41のようなスイッチング構造のための競合の解決を実行するとき、タイムスタンプの値の順序でセルのサービスを提供する（先に列挙した、自己刻時加重式の公正な待ち行列又は仮想時計アルゴリズムのステップ3）ように修正され得るものであると判明した。言い換えれば、予約リング46は、各々のアービトレーション・サイクルの間に存在するいかなる競合をも解決する（即ち、予約リング46の廻りを1周以上伝搬することを変数に要求することなく）。従って、各々の交換器サイクルに関して、ただ1回のアービトレーション・サイクルが要求される。それ故、ルーティング・ネットワーク43又は44のオペレーションに時間差を設けたり、交換に先立って1回のアービトレーション・サイクル以上のアービトレーション・プロセスを開始したりすることは必要でなくなる。その代わり、アービトレーション及び交換のプロセスは、各々のアービトレーション・サイクルの終結において交換サイクル及びもう1回のアービトレーション・サイクルを起動させることによって、直接的に同期され

ることが可能である。

【0031】受け入れられている慣例に従って、有限状態機械(FSM)76a-76nの各々は、スイッチング構造41の入力ポートのそれぞれの1つに関して入力待ち行列の先頭に出現するパケット(HOQパケット)がその他の入力ポートのいずれかのHOQパケットと同じ出力ポート(例えば、出力ポート番号によって)にアドレスされるかどうかを決定するための評価機構として機能する。この目的のため、評価機構76a-76nの各々は、1対のアドレス・レジスタRAx及びSAxを包含する(ここで、xは関連する入力ポートの番号である)。これらのレジスタRAx及びSAxの各々は、関連する入力ポートに関するHOQパケット(もしあれば)の出力ポートアドレスを記憶するために各々のアービトレーション・サイクルの最初に初期化される。入力ポートのいずれかに関する待ち行列がアービトレーション・サイクルの最初からたまたま空であるならば、そのポートにおける評価機構のための状態ロジックは、そのポートに関するビット位置Txを1ビット幅の競合ベクトルに従って真(「1」)状態へセットする。更に、アービトレーション・セッションの始まりの後に到着するパケットが進行中のセッションに参加することを防止するために厳格な「インセッション」制限がアービトレーション・プロセスに賦課されることになるならば、入力ポートxに関する待ち行列がアービトレーション・サイクルの最初から空であるときには必ず、インセッションxフラッグ・ビットが偽(「0」)状態へリセットされることも可能である。

【0032】レジスタRAxにロードされる要求されたアドレスRAは、位置的な変数である。しかし、レジスタSAxにロードされるそのアドレスのコピーSAは、伝搬する変数である。詳細には、アドレスSAは、内部に記憶される要求された位置的なアドレス変数によってその他の評価機構の各々と比較する制御装置78の制御を受けて、頂部から底部への閉ループ順序で予約リング46の廻りをシフトされる。同様に、インセッションxフラッグは、関連する入力ポートがアービトレーション・セッションにおける潜在的な参加者であるかどうかを示す位置的な変数である。他方、競合ベクトルの値Tは、評価機構を介して下向きに伝搬して、問題の評価機構によってサービスされたポートの上か又は下に配置された入力ポートに配置される等しい優先権の競合者の間を評価機構の各々(最上部の評価機構76a以外のもの)に関して識別し、そればかりでなく、現行のアービトレーション・セッションに参加していない入力ポートをも識別する。

【0033】パケット優先権の多数のレベルを評価プロセスに付け加えることも可能である。この目的のため、評価機構76a-76nの各々は、1対のアドレスレジスタRPx及びSPxを包含する(ここで、xは関連す

る入力ポートの番号である)。これらのレジスタRPx及びSPxの各々は、関連する入力ポートに関するHOQパケットの優先権(もしあれば)を記憶するために各々のアービトレーション・サイクルの最初に初期化される。最上位の優先権は優先権レジスタの最小値又はその最大値のいずれかに対して割り当てられることが可能であり；いずれの値が最上位の優先権(即ち、出力への最初のアクセスを付与される)を表現するものであるのかは、以下に説明される優先権比較が比較演算として「小さいか又は等しい」或いは「大きい或又は等しい」のいずれを使用するのかに関する関数である。

【0034】kパケットを越えるものは除いて、kパケットまでのものは、いずれの1回の交換サイクルの間においてもスイッチング構造41の所定の出力ポートへ交換されることが可能である。結果として、評価機構がアービトレーションを失ったときには必ず、公正さは、影響されたパケットがまだアービトレーション・プロセスを受けていないパケットに対して優先権を付与されることを要求する。この理由のため、評価機構76a-76nの各々は、それがアービトレーションを失ったときには必ず、競合フラッグビットFCx及び留保競合フラッグビットFxを真(「1」)状態へセットする。競合留保フラッグビットFxの真(「1」)状態が、不首尾の評価機構に信号を送って、影響されたパケットをその入力待ち行列の先頭に保持するので、進行中のアービトレーション・セッションは、1回又はそれ以上の追加の交換サイクルに渡って延長されることが可能である。他方、競合フラッグビットFCxは、位置的な競合フラッグRCに関する1ビットレジスタRCと、伝搬する競合フラッグSCに関する1ビットレジスタSCxとにコピーされる。評価機構は、参加するか否かに関する決定が為される時点で、その内部における位置的な競合フラッグRCと伝搬する競合フラッグSCとが一致している場合に限って、アービトレーション・セッションに参加する。フラッグRC及びSCは、RP及びSPと共に働き、1回又はそれ以上の交換サイクルの過程に渡って、所定の出力ポートを求めるすべての競合者に優先化された頂部から底部へのラウンドロビン順序でその出力へのアクセスを付与する。

【0035】評価機構76a-76nの各々は、予め定められた開始値からk-1までの範囲に渡って計数する整数カウンタCNTRxを包含する。このカウンタは、各々のアービトレーション・サイクルの最初にその開始値(例えば0)へクリアされ、その後、その入力ポートにおけるパケットがより上位の入力ポートにおけるパケットと競合していることをそのホスト評価機構が決定したときには必ず、増分される。評価機構がその位置的なアドレス変数RAに等しいシフトされたアドレスSAを受け取り、その位置的な優先権変数RPの値がシフトされた優先権SPより小さいか又は等しい(大きい或又は

等しい) 場合には、評価機構は、そのときのその現行の競合ベクトルビットTの状態とその留保競合フラッグFの状態をチェックする。評価機構が真(「1」)の競合ベクトルビットTを有し、その留保競合フラッグFがまだ偽(「0」)であるならば、評価機構は、それがアービトレーションに勝利したものと結論する。従って、アービトレーション・サイクルの終結において、評価機構は、真(「1」)の許諾を、そのCNTRxによって累積された計数(いわゆる「ポート・カウンタ」値)をプラスして、その入力バッファに戻し、それによって、入力待ち行列の先頭におけるセルをそのポートに解放し、仕向け先の出力ポートへ交換することになる。他方、そのCNTRxが偽(「0」)競合ベクトルビット値の存在であふれたことを評価機構が発見したならば、それは最終的にアービトレーションを一巡したものと結論するので、それは、その競合フラッグビットFCx及びその留保競合フラッグビットFxを真(「1」)状態へセットして、上述したように進行中のアービトレーション・セッションを延長する。

【0036】上述のものは、以下のように、より正確に擬似コードで要約されることが可能である：

定数：

$T_0 = 1$  ( $T_0$  はステージ0へのTインプットである)

ポート評価機構についての位置的な変数：

RA: INT [0... ポート数-1] ;

RP: INT [0... 優先権レベル数-1] ;

RC: BOOL ;

FC, F: BOOL ; (F=1=>アービトレーションは喪失)

インセッション: BOOL ;

カウンタ: INT [0... k-1] ;

ポート評価機構についてのシフトされた変数：

SA: INT [0... ポート数-1] ;

SP: INT [0... 優先権レベル数-1] ;

SC: BOOL ;

T: BOOL ; (T=0=>ポートは上にある)

システムリセット：

RA←SA←0 ;

RC←SP←RP←SC←FC←F←T←インセッション←0 ;

カウンタ←0 ;

アービトレーションの開始：

1) パケットがアービトレーションを求めて新たに提出されたなら、FC←0 ;

2) RA←SA←要求されたポート ; SP←RP←要求された優先権 ;

RC←SC←FC ;

F←T←0 ;

インセッション←1 ;

カウンタ←0 ;

3) このポートがこのサイクルを活動中でないならば、

[T←1 ; インセッション←0]

アービトレーション・ステップ：

インセッション=1かつ(RA=SA)ならば、

{

事例：

— 上位の優先権レベルのもう1つのポートが存在する。

SP>RP: カウンタ++; カウンタのあふれにおいて、F←FC←1 ;

— 予約リング内においてこの上に、同じ優先権を有し同じ出力を欲するもう1つのポートが存在する。

(SC=RC)かつ(SP=RP)かつ(T=0): カウンタ++ ;

カウンタのあふれにおいて、F←FC←1 ;

— 明白な競合者が、予約リング内におけるこの下のポート、即ち非活動ポートにあって、同じ優先権を有している。

(SC=RC)かつ(SP=RP)かつ(T=1): 空白 ;

— 明白な競合者が下位の優先権を有する、それは無視せよ。

SP<RP: 空白

— 明白な競合者がこのアービトレーション・セッションの一部ではない。

(RC=1)かつ(SC=0)かつ(SP=RP): 空白 ;

— このポートはこのアービトレーション・セッションにおける有効な参加者ではない。出ていけ！

(RC=0)かつ(SC=1)かつ(SP=RP):

{F←1 ; FC←0 ; インセッション←0 ;} ;

事例終了 ;

};

仕上げ：

FC←インセッション及びF ;

[F=0かつインセッション=1]ならば、パケット(許諾=1)を送れ。

【0037】所望されるならば、前述の予約リング機構46に関するインセッション制限は、それが厳密に現行のセッションの中にあるか否かに関わらず、有効なアービトレーション要求がアービトレーション・セッションに参加することを許容するように緩和されても良い。このより緩和されたアプローチの利点は、空の交換器サイクルの個数を低減させ、それによって交換器の処理能力を増大させることである。この緩和されたバージョンに関する擬似コード記述は、以下の通りである：

定数：

$T_0 = 1$

ポート評価機構についての位置的な変数：

RA: INT [0... ポート数-1] ;

RP: INT [0... 優先権レベル数-1];  
 RC: BOOL;  
 SP: INT [0... 優先権レベル数-1];  
 FC, F: BOOL; (F=1=>アービトレーション  
 は喪失)  
 インセッション: BOOL;  
 カウンタ: INT [0... k-1];  
 アクティブ: BOOL  
 ポート評価機構についてのシフトされた変数:  
 SA: INT [0... ポート数-1];  
 SP: INT [0... 優先権レベル数-1];  
 SC: BOOL;  
 T: BOOL; (T=0=>ポートは上にある)  
 システムリセット:  
 RA←SA←0;  
 RC←SC←FC←F←T←インセッション←0;  
 カウンタ←0;  
 アクティブ←0;  
 アービトレーションの開始:  
 1) パケットがアービトレーションを求めて新たに提出 20  
 されたなら、[FC←0; アクティブ←1;]  
 2) RA←SA←要求されたポート;  
 SP←RP←要求された優先権;  
 RC←SC←FC;  
 F←T←0;  
 インセッション←1;  
 カウンタ←0;  
 3) このポートがこのサイクルを活動中でないならば、  
 T←1;  
 アービトレーション・ステップ:  
 (RA=SA)ならば、  
 {  
 事例:  
 — 上位の優先権レベルのもう1つのポートが存在す  
 る。  
 SP>RP: カウンタ++; カウンタのあふれにおい  
 て、F←FC←1;  
 — 予約リング内においてこの上に、同じ優先権を有  
 し同じ出力を欲するもう1つのポートが存在する。  
 (SC=RC)かつ(SP=RP)かつ(T=0): カ 40  
 ウンタ++;  
 カウンタのあふれにおいて、F←FC←1;  
 — 明白な競合者が、予約リング内におけるこの下の  
 ポート、即ち非活動ポートにあって、同じ優先権を有し  
 ている。  
 (SC=RC)かつ(SP=RP)かつ(T=1): 空  
 白;  
 — 明白な競合者が下位の優先権を有する、それは無  
 視せよ。  
 SP<RP: 空白

— 明白な競合者がこのアービトレーション・セッシ  
 ョンの一部ではない。  
 (RC=1)かつ(SC=0)かつ(SP=RP): 空  
 白;  
 — このポートはこのアービトレーション・セッシ  
 ョンにおける参加者ではない。このセッション内にk以上  
 の有効な参加者が存在するならばチェックせよ。  
 (RC=0)かつ(SC=1)かつ(SP=RP):  
 {インセッション←0; カウンタ++; カウンタのあふ  
 10 れにおいて、F←1}; 事例終了;  
 FC←インセッション及びF (a) セッションに  
 おける参加者の#、プラス(b)セッションの一部では  
 ないがアービトレーションされた出力ポートを要求して  
 いて評価機構の上にある入力ポートの#の和がkより大  
 きいならば、Fは「1」へセットされたことに留意せ  
 よ)。  
 };  
 仕上げ:  
 F=0かつアクティブ=1ならば、ポートカウンタにお  
 けるパケット(許諾=1)を送れ;  
 一旦、許諾が送られてしまうと、アクティブ←0。  
 【0038】E. 入力及び出力緩衝方式—待ち行列戦略  
 スイッチング機構41は、すべての入力線及び出力線  
 の間で待ち行列資源を共用するのではなく、分散された待  
 ち行列、即ち、各々の入力線及び出力線に関して1つ又  
 はそれ以上の専用の待ち行列、を使用する。周知のよう  
 に、共用される待ち行列は、すべての線を横断するバッ  
 ファ空間のダイナミック再配分のために有益である;  
 公正さの保証と帯域幅の予約が要求されるときには増大  
 30 した実現の複雑性を犠牲にして緩衝記憶装置をより効率  
 的に使用する傾向があるアプローチである。しかし、こ  
 の場合は、設計の明快さがバッファ利用の最適化よりも  
 重要であり、交換器に公正なサービス秩序を賦課しつ  
 つ、入力線及び出力線を互いに絶縁するための装備が形  
 成される。結果として、数個の入力がトラフィックを担  
 持する出力の能力を実質的に超過する速度で単一の出力  
 に伝搬するならば、セルはそれらの入力線のためのバッ  
 ファか又はその特定の出力のためのバッファのいずれか  
 において交換器によって失われてしまうことになる。公  
 正さは、入力と出力の両者の待ち行列ポイントにおいて  
 考慮されなければならない。  
 【0039】ネットワーク21(図1)は、保証された  
 トラフィックが接続部のトラフィック制限によって規定  
 される速度を超過しないことをチェックする直接的なト  
 ークン・パケット機構を適切に採用している。詳細に  
 は、ネットワークによってサービスされるコンピュータ  
 のネットワーク・インターフェースのハードウェア/ソ  
 フトウェア(図示略)が、保証されたトラフィックの速  
 度をそのトラフィック制限によって規定されるものに限  
 50 定するのである。標準的な慣例に従って、VCIテーブ

ル内におけるフィールドは、トラフィックの異なった集合の間の差異を識別する。図示された実施例では、VCIテーブル内におけるトラフィック予約(ResTraf)ビット(交換器ヘッダにコピーされる)が、前もって資源の予約を為した保証されたトラフィック・ストリームを表示するために使用される。従って、各々の交換器22-25の各々の入力ポートに関しては、少なくとも2つの入力待ち行列、保証されたトラフィックのための1つと保証されないトラフィックのためのもう1つが存在する。トラフィック制限に従う保証されたトラフィックは、自己刻時加重式の公正な待ち行列又は仮想時計のいずれかによって規定される順序でサービスされ、セル遅延変動(CDV)に大きな制限がある保証なしのトラフィックに対して優先権を付与される。

【0040】保証なしのトラフィックに関して、交換器は、トラフィックをスケジュールするために自己刻時加重式の公正な待ち行列を適切に採用している。保証なしのトラフィックは、供給源が送信する速度を制御するためにネットワークからのフィードバックが使用される閉ループ制御(輻輳回避制御ループ)を採用する。交換器が自己刻時加重式の公正な待ち行列によって規定される順序でセルをスケジュールするとき、ネットワーク安定性のような特性が立証され得ることは周知である。

【0041】出力待ち行列において、次のセルを選択するために必要とされるすべての情報は、位置的に利用可能である。しかし、本発明に従えば、入力待ち行列ポイントにおいて、競合者は、多数の評価装置を横断して展開される。自己刻時加重式の公正な待ち行列のステップを実現するため、各々の評価装置は、要求された出力ポートに関連するシステム仮想時間を知ることが必要であり、同じ出力を要求する入力評価装置は、それらの関連する入力待ち行列の先頭におけるパケットに関連する仮想時間の順序でサービスされなければならない。同様に、仮想時計の場合、仮想時間に関する仮想時計アルゴリズムの計算は評価装置に対して位置的であるデータのみに依存するが、多数の評価装置からのパケットは、それらのパケットが同じ出力に仕向けられているならば、それらの仮想時間の順序でサービスされなければならない。

【0042】優先権を備えた予約リングは、自己刻時加重式の公正な待ち行列のために必要とされるシステム仮想時間と通信するために、更に各々の評価装置の関連する待ち行列の先頭におけるパケットの仮想時間のベースで評価装置を優先化するためにも必要とされる相関性を提供することが可能であると判明している。鍵となる理解は、優先権ビットの個数が十分に大きくなると、優先権とタイムスタンプの間には僅かしか違いがないということである。各々の評価機構がその待ち行列の先頭におけるパケットに関する仮想終了時間を優先権フィールドに入力すると、優先権が増加された予約リングは最下

位の仮想終了時間を備えたkパケットを選択することになる。このステップは、仮想時計と自己刻時加重式の公正な待ち行列の両者におけるソーティング・ステップと同等である。すべての評価機構がすべての要求をアービトレーション・プロセスの間に見るので、いずれの評価機構もが、特定の出力に関連するシステム仮想時間(=先のサイクルにおいてその出力へのアクセスを許諾された最大仮想時間)を容易に発見することが可能であり、それによって自己刻時加重式の公正な待ち行列のその他の構成要素をも可能とするのである。

【0043】優先権ビットの個数が限定されているとすれば、優先権へのタイムスタンプのマッピングは、タイムスタンプが利用可能な優先権の個数よりも大きくなることを理解しなければならない。実現可能なタイムスタンプの範囲(最小から最大まで)が限定されているとすれば、TCPデータ通信プロトコルにおけるシーケンス数の循環のために使用される周知の技術が良好に機能する。特に、その実現可能な値が最大の実現可能な優先権値の半分以下だけ異なる2つの優先権値がa及びbであるならば、2の補数計算が正しい答えを与える。

【0044】我々は、優先権空間を最大の相違の2倍に形成し、即ち1ビット大きくすることによって、実現可能な値が最大の実現可能な優先権の半分以下だけ異なることを保証することができる。

【0045】そのとき、各々の出力ポートに関しては、t最初がサービスされるべき次のパケットの時間であり、t最終がサービスされることになる最後のパケットに関連する時間である、2つの関連する時間： t最初及びt最終が存在する。t最初とt最終の間の相違は、bが優先権ビットの個数である場合に2bの半分より大きいものではあり得ない。これは、その時間をモジュロ2bに維持されるものとして見ることによって保証されることが可能であり、更に続いて、いかなるオフセット(仮想時計又は加重される公正な待ち行列のそれぞれにおける時間速度又は時間加重のいずれかのパケット長さ)も2b-1より大きくないことを保証する。OC-3速度(149.76mbps-SONETペイロード速度)で稼働するATMリンクの場合、このリンクには、ほぼ353208セル/秒が存在する。64Kbps(音声電話速度)回路(AALタイプ1が使用されるときは174セル/秒)が維持されることを要する最低の速度であるならば、最低速度に対する最高速度の比率は2029、即ち11ビットである。この比率は、仮想時間の計算の間に付け加えられることになる最大のオフセットである。従って、12ビットの優先権フィールドは、速度が64Kbpsからフルライン速度にまで及び回路に関連する仮想時間をコード化するために十分なものである。

【0046】入力/出力待ち行列式交換器の場合、入力待ち行列は、スピードアップ要因kによって乗算される

同等のライン速度を経験する。典型的には、入力において待ち行列に入れられるセルの個数は、出力において待ち行列に入れられるものよりも遥かに小さく、従って、入力待ち行列におけるより少ない待機時間という結果になる。我々は、優先権ビットを僅かしか要求しないが最悪でもジッタのための限定的な成果を産み出すことになるソーティング・ステップに対しての近似的な解決を説明する。項目の収集物がキー空間の部分集合に基づいてそれらをバケットへ分割することによってソーティングされ得ることは周知である。完全なソーティングのためには、この分割プロセスは、バケットにおいて再帰的に実行されることを必要とする。この形態におけるソーティング・プロセスは、基数法分類として知られている。レコードが完全にソーティングされることが必要とされないならば、再帰的な基数法分類は、ほぼ同一の値であるがソーティングされた順序ではない項目をバケットに包含させておいて、完了の前に終了されることも可能である。一般的に、仮想的なバケット・スケジューリング時間の近似的な順序付けは、サービス保証に関する従来のな等級を満足させるに足る十分なものである。

【0047】維持されることを要するタイムスタンプの範囲は、仮想スケジューリング時間の計算が入力待ち行列において所定のVCのために現在待機している第1のセル即ちバケットに関してのみ為されることを必要としていることを、注目することによって更に制限され得る。これは、これがそうなることを以下の通りに論証する： 仮想時計の場合、仮想スケジューリング時間は、既に待機しているこのVCのためのセルが存在しないならば、接続仮想時間（VCI変換テーブルからの）と到着時間の関数である。待機しているセルが存在するならば、問題のセルの直前を先行するセルの仮想スケジューリング時間のみが関係する。リンクが過剰に予約を申し込まれて（予約されたトラフィックに関しては不可能）いない限り、仮想時間は、リアルタイムよりも大きいか又は等しくなければならない。同様に、加重される公正な待ち行列の場合も、多数のセルが単一のVCのために待機しているとき、システム仮想時間は現在送信されているセルの仮想時間にセットされるので、セルのための仮想スケジューリング時間は、到着時間よりも大きいか又は等しくなければならない。従って、それは、待ち行列ポイントにおいて現在待機しているVCの最も古いセルの仮想スケジューリング時間のみを維持することで十分なのである。

【0048】自己刻時加重式の公正な待ち行列の場合、各々の入力ポートは、各々の出力に関連するシステム仮想時間を維持しなければならない。しかし、幸いなことに、所定の出力へのアクセスを許諾された最大の優先権が、その出力に関連するシステム仮想時間の妥当な近似値である。（この優先権は、バケット・ソーティングの故に、即ち多数の時間が1つのバケットにマッピングさ

れるので、その仮想時間の推定値である）。従って、各々の入力ポートは、上記の擬似コードで説明された比較によって決定されるように、勝利するポートに関する最大の優先権値が維持される、出力ポート番号によってインデックスされる配列を維持する。このようにして、この優先権値は、優先権値への仮想時間のマッピングを決定するためと、自己刻時式の公正な待ち行列のアルゴリズムのためのシステム仮想時間を導き出すための両方に使用される。

10 【0049】仮想時計の場合、接続仮想時間は、セル到着時間と先の仮想スケジューリング時間のみに依存する。我々は、セル到着時間（インターフェースに対して位置的である）と先の仮想スケジューリング時間（VC情報についての一部として記憶される）を比較し、VCの割当て速度によって決定される定数値を付け加えることによって、接続の仮想時間を計算する。優先権値は、続いて、仮想スケジューリング時間割る2（基数）を取り、関連する出力へのアクセスを先に許諾された最大の優先権を付け加えることによって計算される。

20 【0050】自己刻時式の公正な待ち行列の場合、システム仮想時間は、接続の仮想スケジューリング時間を計算するように要求される。我々は、その出力ポートに関するシステム仮想時間の推定値として、関連する出力へのアクセスを先に許諾された最大の優先権掛ける2（基数）を使用する。VCについて、VC情報についての一部として記憶される、この接続における先のセルに関する仮想スケジューリング時間が与えられると、アービトレーション・プロセスの残りは、上述したように進行する。

30 【0051】上記の関数は、VC毎のベースで入力待ち行列へ出入りするようにセルを挿入しかつ取り出すための装備が形成されることを要求する。通常の熟練技術者は、各々のリストが、処理されるべき次のセルであるリストの先頭と、このVCに受け取られた次のセルにリンクされることになる末尾とを有する（即ち、VC上をルーティングされる新しいセルは、リンクされたリストに追加される）ようにした、VC毎にリンクされたセルのリストをどのように維持するのか理解するであろう。

40 【0052】代替的に、入力／出力待ち行列式交換器の目的のためには、完全にリンクされたリスト構造を維持することは過剰であるかもしれない。これは2つの理由で正しい： 第1に、我々は、VCリストについてその先頭にある1つのセル（最初の仮想スケジューリング時間）のみを必要とし、第2に、入力待ち行列の寸法は、おそらく小さいものであることになるからである。完全にリンクされたリストを維持することに対する代替的なアプローチは、入力待ち行列を2つの区域：セルがランダムアクセスのためにロードされ記憶され得る待ち行列の前部、及びFIFO待ち行列として機能する後部に、分割することによって結果を推定することである。限定

された前部区域は、その後、各々のセル時間を走査され、最小の仮想スケジューリング時間によってセルのための線形時間走査を付与する；限定された前部区域の寸法は、区域全体を走査するためにセル時間において十分なクロックサイクルが存在するようにして選択される。各々のセルに対するシーケンス番号の付加及び比較の中への前記シーケンス番号の算入は、仮想回路のセルを順序正しく維持するに足る十分なものである。通常の熟練者は、そのようなアルゴリズムをどのように実現するか理解するであろう。

【0053】予約されたトラフィックに上位の優先権を付与するために、我々は、仮想時間から独立しているログ（優先権等級の数）ビットの優先権フィールドを有している。先に説明した優先権ベースの予約リングをマルチキー・ソーティングに拡大するためには余分のフィールドが使用される。その方法は、当該分野における通常の熟練者には熟知されているであろう。支持される2つの優先権等級が存在する場合、このフィールドは、1ビットであり、SRB及びRRBと呼ばれることになる。

【0054】出力のためのシステム仮想時間を計算するため、各々の入力は、予約リングを監視して、所定の出力に関するk番目の最小の優先権を見つけなければならない（その出力のためにkと同じほど多くの要求が存在するとしたらば）。我々は、各々のアービトレーション・ステップにおいて更新され得るk個の最小値の集合としてこれを実現することが可能である：

最小優先権：INTの配列[0..ポート数-1, 0..k-1]；

更新されたこのサイクル：BOOLの配列[0..ポート数-1, 0..k-1]；——ヘクリアされる——各々のアービトレーション・サイクルの開始において偽（アービトレーションの初期設定）

各々のアービトレーション・ステップにおいて手順を遂行せよ：

k番目の最小優先権(SA, SP)を計算せよ；  
最小優先権の配列及び更新されたこのサイクルの配列を更新するためである。自己刻時加重式の公正な待ち行列が予約と予約なしの両方のトラフィックのために使用されるならば、k番目の最小優先権が予約と予約なしの両方のトラフィック等級に関して計算されなければならない。しかし、我々は、予約されたトラフィック等級に関する仮想時計を使用し、予約されたk番目の最小優先権を計算する必要性を回避することになる。

【0055】k=2の具体例の場合、k番目の最小優先権の計算の手順の具体例実行は：

k番目の最小優先権を計算せよ：手順(SA:INT[0..ポート数-1]), SP:INT[0..優先権レベル数-1]；

{

更新されたこのサイクルが[SA, 0]でないならば、

{最小優先権[SA, 0] ← SP；更新されたこのサイクル[SA, 0] ← 真；} ほかにも、更新されたこのサイクルが[SA, 1]でないならば、

{最小優先権[SA, 1] ← SP；更新されたこのサイクル[SA, 1] ← 真；} ほかにも、SP < 最小優先権[SA, 0]ならば、

最小優先権[SA, 0] ← SP

ほかにも、SP < 最小優先権[SA, 1]ならば、

最小優先権[SA, 1] ← SP；

10 }；——k番目の最小優先権の計算の具体例実行の終了出力のためのシステム仮想時間は、そのとき、以下によって与えられる：

max(最小優先権[j, 0], 最小優先権[j, 1])

各々の評価装置において遂行される予約リング・アルゴリズムは、そのとき：

定数：

T<sub>0</sub> = 1；

予約トラフィック = 1を規定せよ；——最上位の優先権のトラフィック等級

予約なしトラフィック = 0を規定せよ；

トラフィック等級数 = 2を規定せよ；——この具体例の場合、我々は予約及び予約なしのみを支持する。

ポート評価機構についての位置的な変数：

最小優先権：INTの配列[0..ポート数-1, 0..k-1]；

更新されたこのサイクル：BOOL配列[0..ポート数-1, 0..k-1]

；RA:INT[0..ポート数-1]；

30 RP:INT[0..優先権レベル数-1]；

RC:BOOL；

RRB:INT[0..トラフィック等級数-1]；

FC, F:BOOL；(F=1=>アービトレーションは喪失)

インセッション:BOOL；

カウンタ:INT[0..k-1]；

アクティブ:BOOL

ポート評価機構についてのシフトされた変数：

SA:INT[0..ポート数-1]；

40 SP:INT[0..優先権レベル数-1]；

SC:BOOL；

SRB:INT[0..トラフィック等級数-1]；

T:BOOL；(T=0=>ポートは上にある)

システムリセット：

RA ← SA ← 0；

RC ← SC ← FC ← F ← T ← インセッション ← 0；

カウンタ ← 0；

アクティブ ← 0；

最小優先権[\*, \*] ← 0；

50 アービトレーションの開始：

予約された待ち行列にセルが存在するならば、最下位の仮想スケジューリング時間を備えたセルを選択し、予約された待ち行列からそれを取り除き、ほかに、予約なしの待ち行列にセルが存在するならば、最下位の仮想スケジューリング時間を備えたセルを選択し、予約なしの待ち行列からそれを取り除くこと。予約された待ち行列に関する仮想スケジューリング時間は、仮想時計アルゴリズムに従って計算され得る。以下のアルゴリズムにおいて、仮想時計は、各々の出力に関するシステム仮想時間を維持する必要性を回避するために、予約されたトラフィックに関するものと想定されている。予約なしの待ち行列に関する仮想スケジューリング時間は自己刻時加重式の公正な待ち行列に従って計算されるべきであり、出力についての予約なしのトラフィックに関するシステム仮想時間も計算されなければならない。

1) パケットがアービトレーションを求めて新たに提出されたなら、 $[FC \leftarrow 0; \text{アクティブ} \leftarrow 1;]$

2)  $RA \leftarrow SA \leftarrow$  要求されたポート;

$SP \leftarrow RP \leftarrow$  要求された優先権; — 仮想時間から導き出される。

$RC \leftarrow SC \leftarrow FC;$

$F \leftarrow T \leftarrow 0;$

インセッション  $\leftarrow 1;$

カウンタ  $\leftarrow 0;$

更新されたこのサイクル  $[*, *] \leftarrow$  偽;

3) このポートがこのサイクルを活動中でないならば、 $T \leftarrow 1;$

4)  $SRB \leftarrow RRB \leftarrow$  セルが予約されたトラフィックならば予約トラフィック、でなければ予約なしトラフィック;

アービトレーション・ステップ:

$SRB$  でなければ、 $k$  番目の最小優先権 ( $SA, SP$ ) を計算せよ;

( $RA = SA$ ) ならば、

{

事例:

— 上位の優先権トラフィック等級にあるか又はより小さな優先権を備えているいずれかのもう1つのポートが存在する。

( $SRB > RRB$ ) 又は ( $SP < RP$ ): {カウンタ++ ; カウンタのあふれにおいて、 $F \leftarrow 1;$ }

— 我々のアービトレーション・セッションの一部であり、同じ優先権を有し、同じトラフィック等級に所属し、予約リングにおいてこのポートの上にある、もう1つのポートが存在する。

( $SC = RC$ ) かつ ( $SP = RP$ ) かつ ( $RRB = SRB$ ) かつ ( $T = 0$ ):

{カウンタ++ ; カウンタのあふれにおいて、 $F \leftarrow 1;$ }

— 明白な競合者が、スイッチング構造内の下位レベ

ルにおけるポート、即ち非活動ポートにおいて、我々のアービトレーション・セッションの一部であって、我々と同じトラフィック等級及び同じ優先権にある。

( $SC = RC$ ) かつ ( $T = 1$ ) かつ ( $SP = RP$ ) かつ ( $SRB = RRB$ ):

空白;

— 明白な競合者が、我々より下位の優先権トラフィック等級にあるか又は上位の優先権を有する。

( $SRB < RRB$ ) 又は ( $SP > RP$ ): 空白;

10 — 明白な競合者がこのアービトレーション・セッションの一部ではない。

( $RC = 1$ ) かつ ( $SC = 0$ ) かつ ( $SP = RP$ ) かつ ( $SRB = RRB$ ):

空白;

— このポートはこのアービトレーション・セッションにおける参加者ではない。このセッション内に  $k$  以上の有効な参加者が存在するならばチェックせよ。( $RC = 0$ ) かつ ( $SC = 1$ ) かつ ( $SP = RP$ ) かつ ( $SRB = RRB$ ):

20 {インセッション  $\leftarrow 0$ ; カウンタ++ ; カウンタのあふれにおいて、 $F \leftarrow 1$ };

事例終了;

$FC \leftarrow$  インセッション及び  $F$  ((a) セッションにおける参加者の#, プラス (b) セッションの一部ではないがアービトレーションされた出力ポートを要求していて評価機構の上にある入力ポートの#の和が  $k$  より大きいならば、 $F$  は「1」へセットされたことに留意せよ)。

} ;

30 仕上げ:

$F = 0$  かつ  $\text{アクティブ} = 1$  ならば、ポートカウンタにおけるパケット (許諾 = 1) を送れ;

一旦、許諾が送られてしまうと、 $\text{アクティブ} \leftarrow 0$ 。

【0056】図8は、その2つのパニヤン・ネットワーク43及び44を図2で示されたように備えた構造41のように、2つのルーティング・ネットワーク ( $k = 2$ ) を有する自動ルーティング・非ブロッキング式スイッチング構造に関して、参照番号46で示されたような形式の予約リングによって実行されるアービトレーションを概略的に示している。この簡略化された具体例では、スイッチング構造は、パケットヘッダに包含される出力ポートアドレス即ち宛先をベースとして、4つの入力ポート、 $InPort0-InPort3$ からのパケットを、4つの出力ポート、 $OutPort0-OutPort3$ の中の指定されたものへルーティングする。予約リングは、各々のアービトレーション・サイクルの間に、スイッチング構造の入力ポートのすべてをアービトレーションして、同じ出力ポートにアドレスされたパケットの間において発生するいかなる競合をも解決する。競合を完全に解決するためにマルチサイクル・アー



ビットレシジョン・セッションが求められるかもしれないが、k個までの競合するパケットのために競合が解決されるときには、アービトレシジョン・セッションの最後のサイクルを例外として（即ち残留する競合者）、競合は、k個の競合するパケットのために各々のアービトレシジョン・サイクルにおいて解決される。従って、予約リングは、アービトレシジョンされたパケットを各々のアービトレシジョン・サイクルの終結時にスイッチング構造の中へ解放し、各々のアービトレシジョン・サイクルの直後には交換サイクルが随伴される。

【0057】より詳細には、アービトレシジョン・セッションは、OutPort3にアドレスされるInPort0、InPort1、及びInPort3のためのHOQパケットによって、更にはOutPort0にアドレスされるInPort0のためのHOQパケットによって開始されるかもしれない。これは新しいアービトレシジョン・セッションであると想定されることになる（即ち、いずれのパケットも進行中のアービトレシジョン・セッションを受けていない）。従って、予約リングの評価機構は、すべてが、アービトレシジョン・セッションのためにそれらの状態制御ロジックによって初期化される。この初期設定プロセスの結果として、スイッチング構造の入力ポートの各々に関する評価機構は、その特定の入力ポートにおけるHOQパケットがその要求されたアドレスレジスタRAxの中へ更にそのシフトされたアドレスレジスタSAxの中へルーティングされコピーされることになる、出力ポートのアドレスを有する。更に、各々の評価機構は、その要求された優先権RPx及びシフトされた優先権SPxのレジスタにパケットの優先権（仮想タイムスタンプに等しい）をロードし、その要求された予約ビットRRBx及びシフトされた予約ビットSRBxにパケットのトラフィック等級をロードする。入力ポートのすべては、このアービトレシジョン・セッションの間たまたま活動的であるので、各々の評価機構のインセッション・ビットは、真（「1」）状態へセットされる。更になお、これが新しいアービトレシジョン・セッションであるので、各々の評価機構は、そのカウンタCNTxをクリアさせ、その競合に関係するすべてのビットF、FC、RC及びSCを、その競合ベクトルビットTと共に偽（「0」）状態へリセットする。

【0058】アービトレシジョン・プロセスのステップ0において、各々の評価機構は、そのシフトされたアドレスレジスタSAxに記憶されたアドレス、そのシフトされた競合フラッグビットSC、そのシフトされた優先権SPx、及びそのシフトされた予約ビットSRBxを、下方へ閉ループ・ラウンドロビン順序で、予約リングの次の下位の評価機構の中へシフトする。これが閉ループ・シフトであるので、最下位の評価機構（即ちInPort3のための評価機構）に関するシフトされたア

ドレス、シフトされた競合フラッグビット、シフトされた優先権及びシフトされた予約ビットは、最上位の評価機構（即ちInPort0のための評価機構）の中へシフトされる。同時に、真（「1」）の競合ベクトルビットはその最上位の評価機構の中へシフトされ、既存の競合ベクトルビットは次の下位の隣接する評価機構の中へ閉ループ様式で下方にシフトされる（即ち、最下位の評価機構からの競合ベクトルビットは落される）。

【0059】シフトオペレーションに続いて、各々の評価機構は、その要求されたアドレスレジスタRAx内のアドレスを、そのシフトされたアドレスレジスタSAx内のアドレスとの同等性で比較する。それらの2つのアドレスが異なっていると評価機構が決定するならば、それはこれ以上の処置を取らない（InPort2及びInPort3に関する評価機構の状態を参照すること）。他方、その要求されたアドレスとその中へシフトされたアドレスとが同じであると評価機構が決定するならば、それは、その入力ポートにおけるパケットがもう1つの入力ポートにおけるパケットと競合しているかもしれないと結論する。産物の優先権値及びトラフィック形式は、競合するパケットのレベルを増大させる。我々は、RPx及びSPx及びRRBx及びSRBxを比較して、シフトされた要求が下位であるか、上位であるか、それとも同じ優先権であるかを決定する。優先権が同じであれば、それらのパケットの間に実際的な競合が存在することを確認するために、評価機構は、その要求された競合ビットRC及びシフトされた競合ビットSCの状態を同等性に関して比較する。それらの2つのビットが同じ状態（両方ともが偽（「0」）又は真（「1」））を有するならば、評価機構は、その競合ベクトルビットTの状態をチェックして、評価機構がサービスしている入力ポートの上又は下のいずれの入力ポートに競合するパケットが存在しているのかを決定する。競合者が下位の入力ポートにあるならば（真（「1」）状態にある競合ベクトルビット）、評価機構に直接的な処置は何も要求されない。他方、競合者が上位の入力ポートにあると評価機構が結論するならば（まだ偽（「0」）状態にある競合ベクトルビット）、評価機構は、そのカウンタCNTxを増分し、競合が解決されたという事実を競合者に報告する。

【0060】このステップ及び比較プロセスは、各々のアービトレシジョン・サイクルに関して少なくともN-1回（ここで、Nは入力ポートの数である）繰り返される。いずれかの評価機構のカウンタCNTxがそのプロセスのいずれかのステップの間にあふれるならば、その評価機構に関する留保競合フラッグビットFは、真（「1」）状態へセットされ、それによって、その評価機構の競合フラッグビットFCをも真（「1」）状態へセットすることになる。

【0061】所望されるならば、アービトレシジョン・

サイクルの終結の信号を送るために補足的なステップ・オペレーションが実行されても構わない。アービトレーション・サイクルの終結において、評価機構は、それらの留保競合ビットFがまだ偽である場合に限り、真（「1」）許諾信号をそれらのそれぞれの入力ポートに返還する。そのような許諾を受け取ったいずれの入力ポートも、そのH O Qパケットをスイッチング構造の中へ解放し、指定された出力ポートへルーティングさせる。理解されるであろうが、各々の評価機構は、その留保競合ビットFをもその関連する入力ポートに返還する。従って、入力ポートが偽（「0」）の許諾及び真（「1」）の留保状態ビットを同時に受け取るならば、入力ポートは、その現行のH O Qパケットを次のアービトレーション・サイクルのために保持する。

【0062】アービトレーション・セッションの第2及び後続のアービトレーション・サイクルは、第1のものと同様である。即ち、評価機構内に記憶される要求されたアドレス及び要求された競合フラッグビットは、アドレス及び競合状態データの2次元配列を提供し、これに対して、循環するシフトされたアドレス及びシフトされた競合フラッグビットによって提供されるアドレス及び状態データが入力ポートのあらゆる対に関して同等性を比較されるのである。しかし、アービトレーション・セッションを完了させるために補足的なアービトレーション・サイクルが要求されるときには必ず、そのカウンタC N T Rをあふれさせた（図8を参照）いずれかの評価機構における先行するアービトレーション・サイクルの完了に先立って、競合フラッグビットF Cが真（「1」）状態へセットされる。このフラッグビットF Cは、続いて、次のアービトレーション・サイクルのためのその初期設定（図を参照）の間に、その評価機構の要求された競合ビットR C及びシフトされた競合ビットS Cの中へコピーされ、それによって、未解決の競合がアービトレーション・セッションの後続のサイクルの間に頂部から底部へのラウンドロビン順序で解決され続けることを保証することになる。

【0063】要求されたアドレスとその中へシフトされたアドレスの間に同等性が存在することを発見する評価機構は、それが、シフトされたアドレスには真

（「1」）のシフトされた競合フラッグビットS Cが同伴されていることと、それ自体の要求された競合フラッグビットR Cが偽であることを発見するならば、そのインセッション・ビットを偽（「0」）状態へリセットする。

【0064】本発明はパッチャ／パニヤンATMスイッチング構造のためのアービトレーション・プロセスを特に参照して説明されてきたが、それがクロスバー・アーキテクチャを有するもののようなATMスイッチング構造のその他の形式のためのアービタにおいて採用され得

ることは明白であろう。幾つかのATMスイッチング用途に関しては、クロスバー・アーキテクチャの交換器構造が好適である。

【0065】更に一般的に、熟練技能者は、本発明の多くの機構がIP（インターネット・プロトコル）ネットワーク及びフレームリレー・ネットワークのようなパケット交換式ネットワークのその他の形式に関するアービトレーション分散型多重化に対して適用可能であることを認識するであろう。従って、「アービトレーション分散型多重化ポイント」を、アービタの制御を受けて1つ又はそれ以上の出力に対して複数の入力多重化されたアクセスを提供し、このアービタが、続いて、入力を横断して分散される入力待ち行列の中に存在するアクセス要求の間をアービトレーションするようにしたシステム又はサブシステム（本文では、集散的に「システム」と呼ばれた）であると定義することが適切である。

#### 【図面の簡単な説明】

【図1】 階層的なATMネットワークを概略的に示している。

【図2】 ATM交換器を概略的に示している。

【図3】 AはATMセルのための通常フォーマットを示している。BはATMセルの典型的な交換器ヘッダを示している。

【図4】 パッチャ・ソーティング要素を概略的に示している。

【図5】 パッチャ・ソーティング要素のオペレーションを示す有限状態機械である。

【図6】 図4で示された形式のソーティング要素の再帰的な組合せから構成されるパッチャ・ソーティング・ネットワークを概略的に示している。

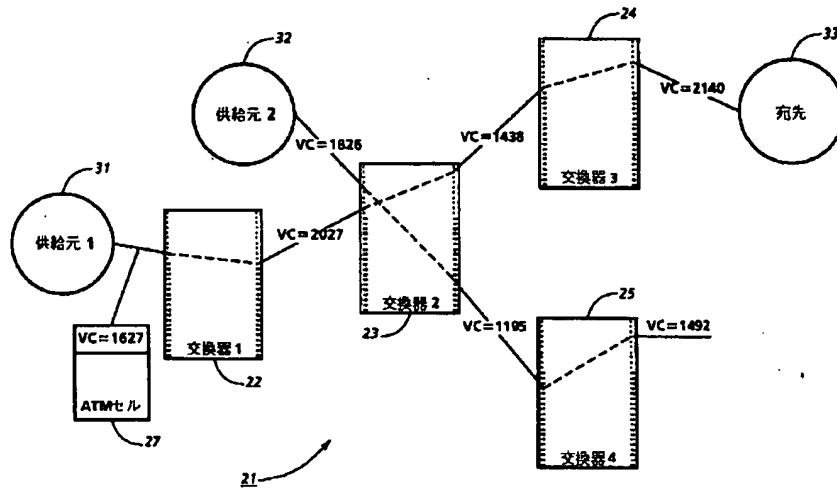
【図7】 パニヤン・ルーティング・ネットワークを概略的に示している。

【図8】 本発明を実行するものとして適切である予約リングの簡略化された概略図である。

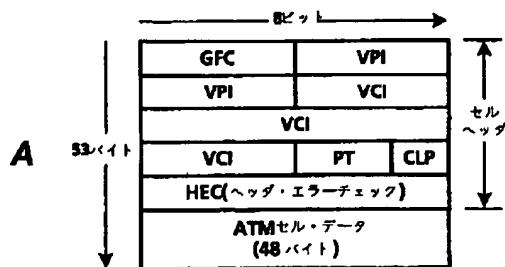
#### 【符号の説明】

21 ネットワーク、22 交換器、23 交換器、24 交換器、25 交換器、27 ATMセル、31 供給元、32 供給元、33 宛先、41 パッチャ／パニヤン・ネットワーク、42 パッチャ・ソーティング・ネットワーク、43 パニヤン・ルーティング交換器、44 パニヤン・ルーティング交換器、45 バッファ、46 予約リング、48 コピーネットワーク、50 記憶装置、51 バッファ、52 バッファ、53 VCI交換機構、54 VCIテーブル参照用記憶装置、62 ソーティング要素、65 パッチャ・ソーティング・ネットワーク、71 パニヤン・ルーティング・ネットワーク、72 スwitching要素、75 線形収縮配列

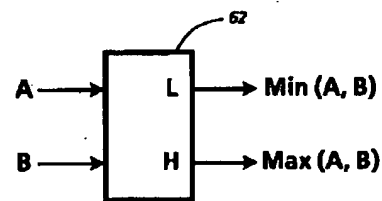
【図1】



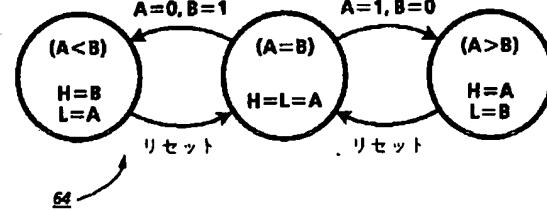
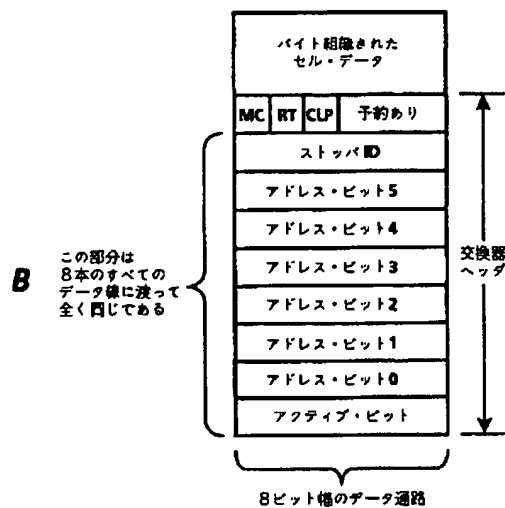
【図3】



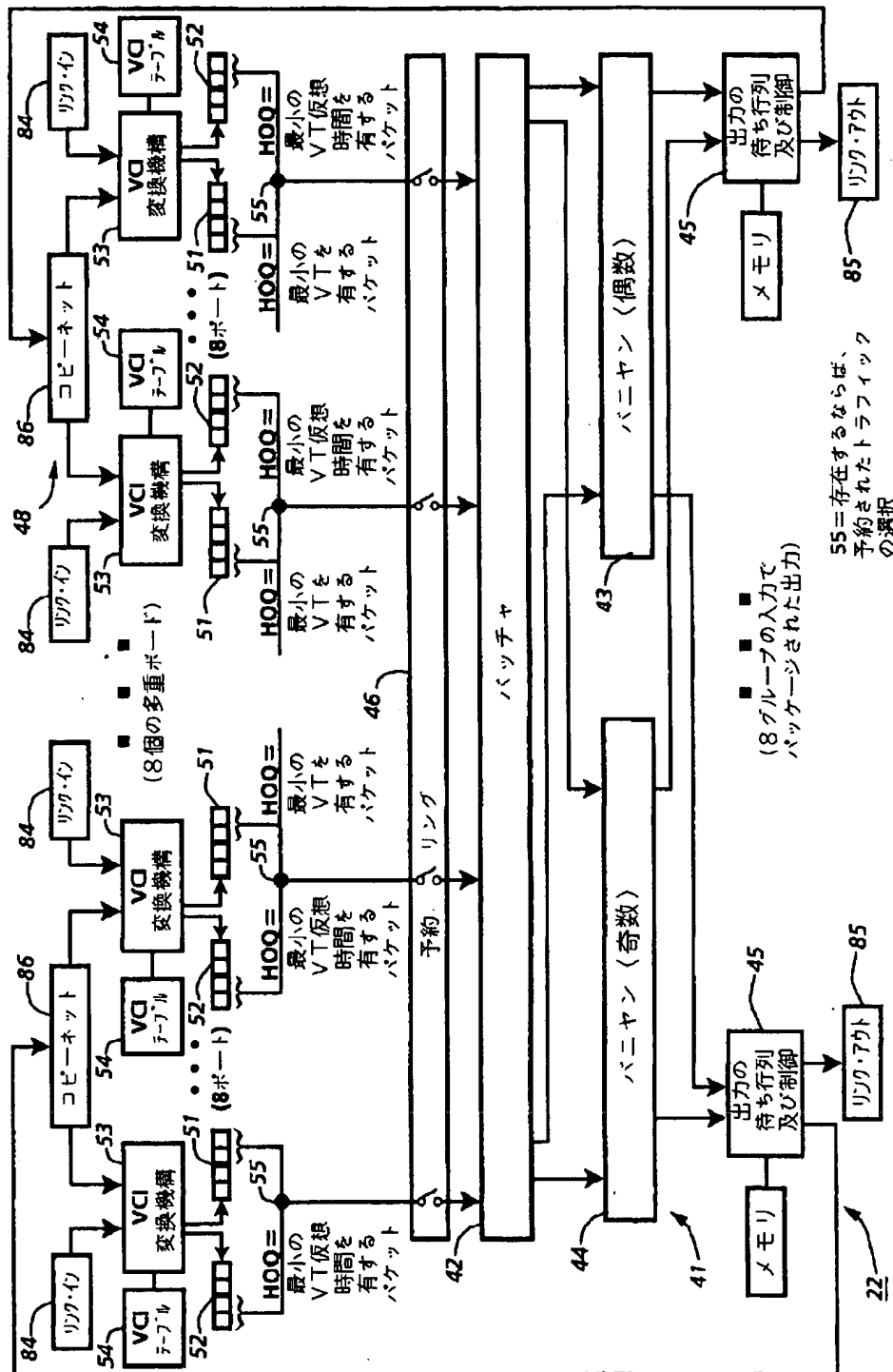
【図4】



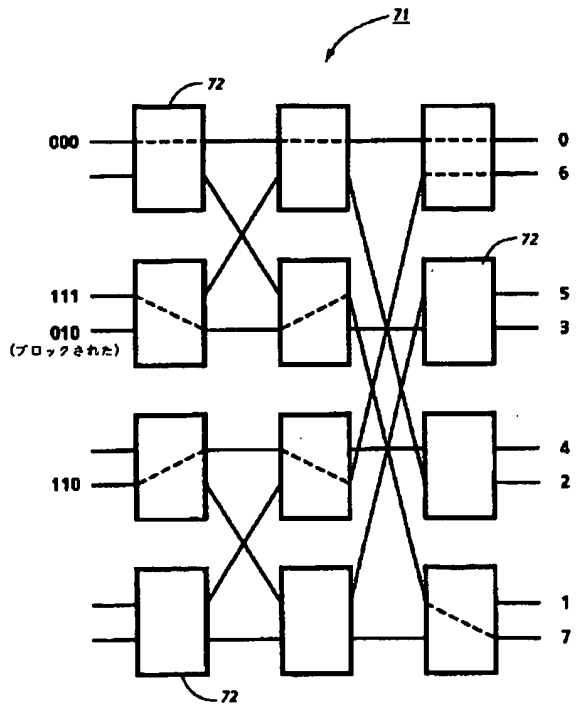
【図5】



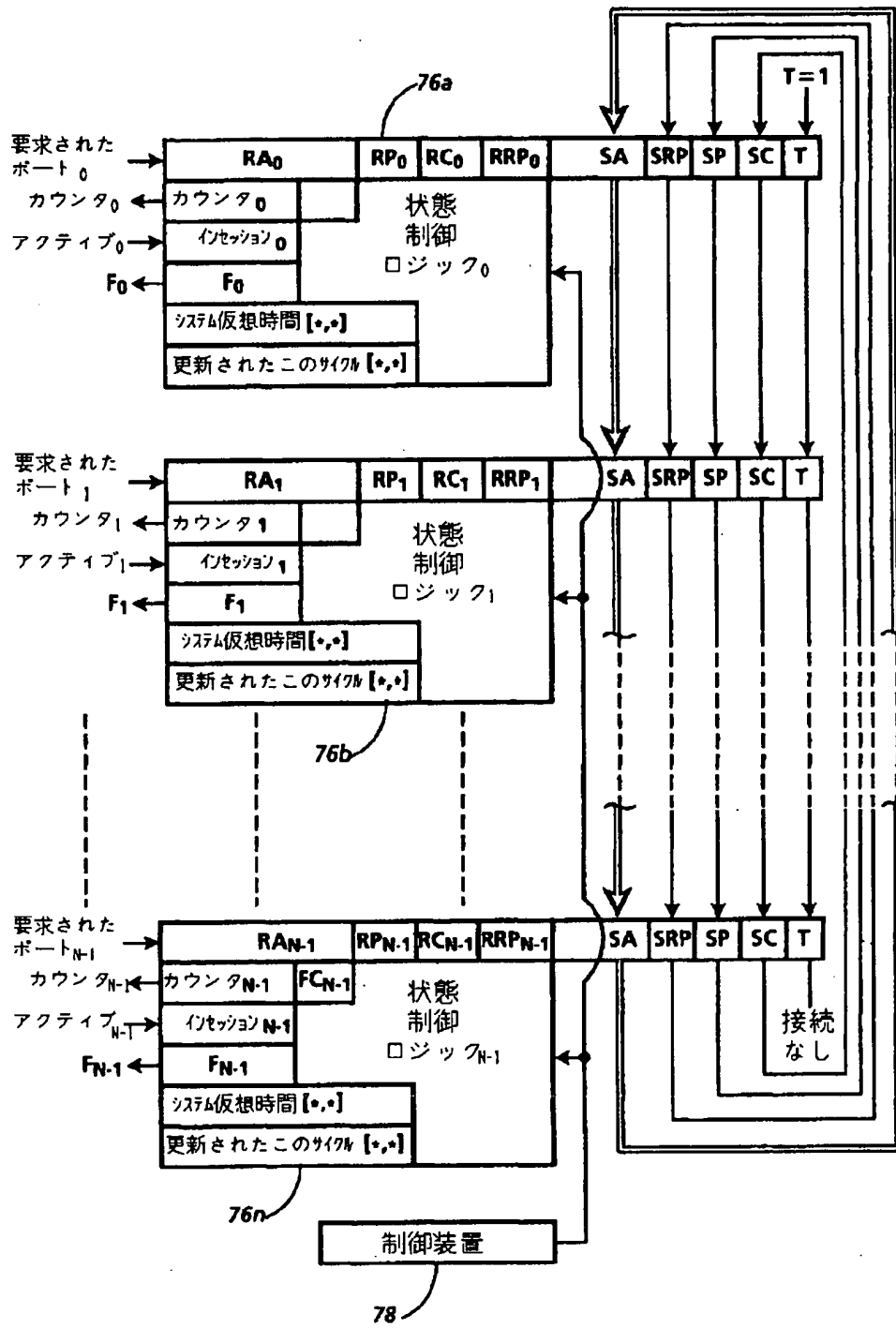
【図2】



【図7】



【図8】



**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record**

**BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

☐ BLACK BORDERS

☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES

☒ FADED TEXT OR DRAWING

☒ BLURRED OR ILLEGIBLE TEXT OR DRAWING

☐ SKEWED/SLANTED IMAGES

☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS

☐ GRAY SCALE DOCUMENTS

☒ LINES OR MARKS ON ORIGINAL DOCUMENT

☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY

☐ OTHER: \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.**